PNAS
www.pnas.org

Supplementary Information for

Mesoscale Modeling Reveals Formation of an Epigenetically Driven HOXC Gene Hub

Gavin D. Bascom, Christopher G. Myers, Tamar Schlick[1]

[1]To whom correspondence should be addressed. E-mail: schlick@nyu.edu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1816424116

Supporting Information Text

1. Supplemental Information

Additional Methods

Model Constituents The histone tails are modeled at the resolution of 5 amino acids per bead, and LHs are represented similarly by 6 beads for the globular head (GH) and 22 beads for the flexible C-terminal Domain (CTD) as described in Ref.(1) Histone tail acetylation is modeled by introducing two discrete states for each tail, an acetylated version of the tail and a wildtype version of the tail, where transitions between these states are controlled by a Metropolis MC step. Essentially, the acetylated tail has a more rigid secondary structure, as determined by all-atom molecular dynamics simulations (2). See Ref. (3, 4) for further model details and validation of equilibrium and dynamic properties.

Starting Configurations: The 3D starting structure for all systems is obtained by generating ideal zigzag conformations oriented with the fiber axis parallel to the z-axis. The z-rise (i.e., distance between successive nucleosomes along the long fiber axis) and fiber width (i.e., distance between successive nucleosomes when viewed in the plane perpendicular to the fiber axis) are assigned values proportional to the DNA linker length between each nucleosome such that DNA/DNA linker bead distances are adjusted to be less than the DNA bead radius (3 nm). Any overlaps between cores or linker DNA beads are removed. Nucleosomes are oriented perpendicular to the central axis, as found to be optimal (5). See Fig. S1 for a visual rendering.

Contact Data: Contact probability matrices describe the fraction of MC steps that any element (i.e., core, tail, or linker DNA) of core i is 'in contact' with any element of core j. A contact is defined when any of these elements are within 2 nm of one another. Contact probabilities are normalized across all frames for a single trajectory and then all probabilities are summed and plotted in logscale. The 1D projection of the contact matrix represented by $I'(i,j)$ is calculated by summing across each row k of the matrix:

$$I(k) = \frac{\sum_{i=1}^{N_C} I'(i, i \pm k)}{\sum_{j=1}^{N_C} I(j)}, \qquad [1]$$

Where N_C refers to the number of cores in the system. Such summation indicates the fraction of configurations between nucleosomes separated by k nucleosomes where a contact is likely (i.e., $i \pm k$ nucleosome neighbors). Fig. S2 provides further detail of the contact map corresponding to the Hox-C system shown in Fig. 2 of the main text.

The contact contributions due to epigenetic elements (main text Fig.6a) were calculated by summing columns in the contact matrix, and counting a contact if both i and j elements contain the given epigenetic feature. All other contacts are considered wildtype. Average pairwise distances (main text Fig. 6b) between promoters are calculated as the Cartesian distance between the center of each promoters +1 nucleosome. Promoters are located at positions: 12:54,378,866 (HOXC10), 12:54,394,097 (HOXC9), 12:54,403,439 (HOXC8), 12:54,410,750 (HOXC6), and 12:54,427,103 (HOXC5). All distances were then averaged across all promoter pairs and across the entire 50 trajectory ensemble. A plot of the average distance was calculated using a bin width of 10 nm. Pairwise distances shown in Fig. 6b are taken from a single trajectory and plotted as a function of simulation length.

Volume/Compaction Measures: The volume of each fiber conformation is calculated using the AlphaShape protocol in Matlab, which creates a convex bounding surface enclosing all elements of the fiber. Surfaces are also visually inspected to ensure that the bounding surface reasonably reflects the fiber morphology. The volume enclosed by the surface is reported as the fiber volume (see Fig. S3.)

The radius of gyration (R_g^2), or the root mean squared distance of each point from the center of mass, was calculated according to the equation:

$$R_g^2 = \frac{1}{N} \sum_{j=1}^{N} (r_j - r_{mean})^2, \qquad [2]$$

where r is the center position of each nucleosome core, and r_{mean} is the average of all core positions, as detailed in Ref. (6)

To calculate the exponent of volume/system size scaling, volume and R_g^2 values were calculated using the above techniques for fragments of fiber conformations with increasing size. Volumes and R_g^2 for each fragment size was calculated for a fragments of the same size across the entire ensemble, and the average is presented for each system size. Volume and R_g^2 were then plotted as a function of fragment size (in base pairs) and fitted to power-law exponentials, where the exponent is presented as the scaling exponent in Table S1.

Additional Results

Uniform fibers show homogenous contacts along the diagonal and at long ranges (Fig. S4a), where **Uniform+NFR** fibers have long-range contacts that are similarly distributed in terms of genome position, but diminished near the diagonal at NFRs (Fig. S4b). **Life-like** fibers similarly show short range contacts inside each clutch (Fig. S5a), but they exhibit fewer contacts at long distances, whereas **Life-like+LH** fibers show contact matrices very similar to life-like fibers but with increased looping between LH/NFR rich regions (Fig. S5b).

To further quantify the scaling behavior of the compaction as a function of system size, we computed the volume and R_g^2 as a function of individual chromatin fragments of varying sizes for **HOXC**, **Life-Like+Ac**, and **Life-Like** fibers. These curves were then fit to a power-law exponential, providing the scaling exponent for volume and R_g^2 function of fiber size (Table S1). We find that HOXC Gene fibers form the most dense structure with the smallest scaling exponents. All volume scaling

exponents are larger than 1, and R_g^2 scaling exponents less than 1, indicative of scaling behavior similar to those found by super-resolution imaging techniques reported by Beottiger et al.(7).

Table S1: Scaling Exponents

System	Volume Scaling Exponent	R_g^2 Scaling Exponent
HOXC Gene	2.03±.01	.48±.04
Uniform+NFR	2.05±.01	.55±.06
Uniform	2.12±.01	.55±.02
Life-Like+LH	2.14±.01	.67±.03
Life-Like+Ac	2.14±.01	.57±.02
Life-Like	2.88±.01	.64±.02

The sedimentation coefficient (denoted $S_{20,w}$) is the ratio of a components velocity to the acceleration that it experiences, given in the units of Svedbergs. It can be approximated from S_{N_C} where

$$\frac{S_{N_C}}{S_1} = 1 + \frac{R_1}{N_C} \sum_i \sum_j \frac{1}{R_{ij}}. \quad [3]$$

Here, S_{N_C} represents the sedimentation coefficient $S_{20,w}$ for a rigid structure consisting of N_C nucleosomes of radius R_1. R_{ij} represents the distance between the center of two nucleosomes i and j, and S_1 is the sedimentation coefficient for a mononucleosome. We use R_1=5.5 nm, and $S_1 = 11.1$ Svedbergs (1S = 10^{-13} s), as detailed in Ref. (6).

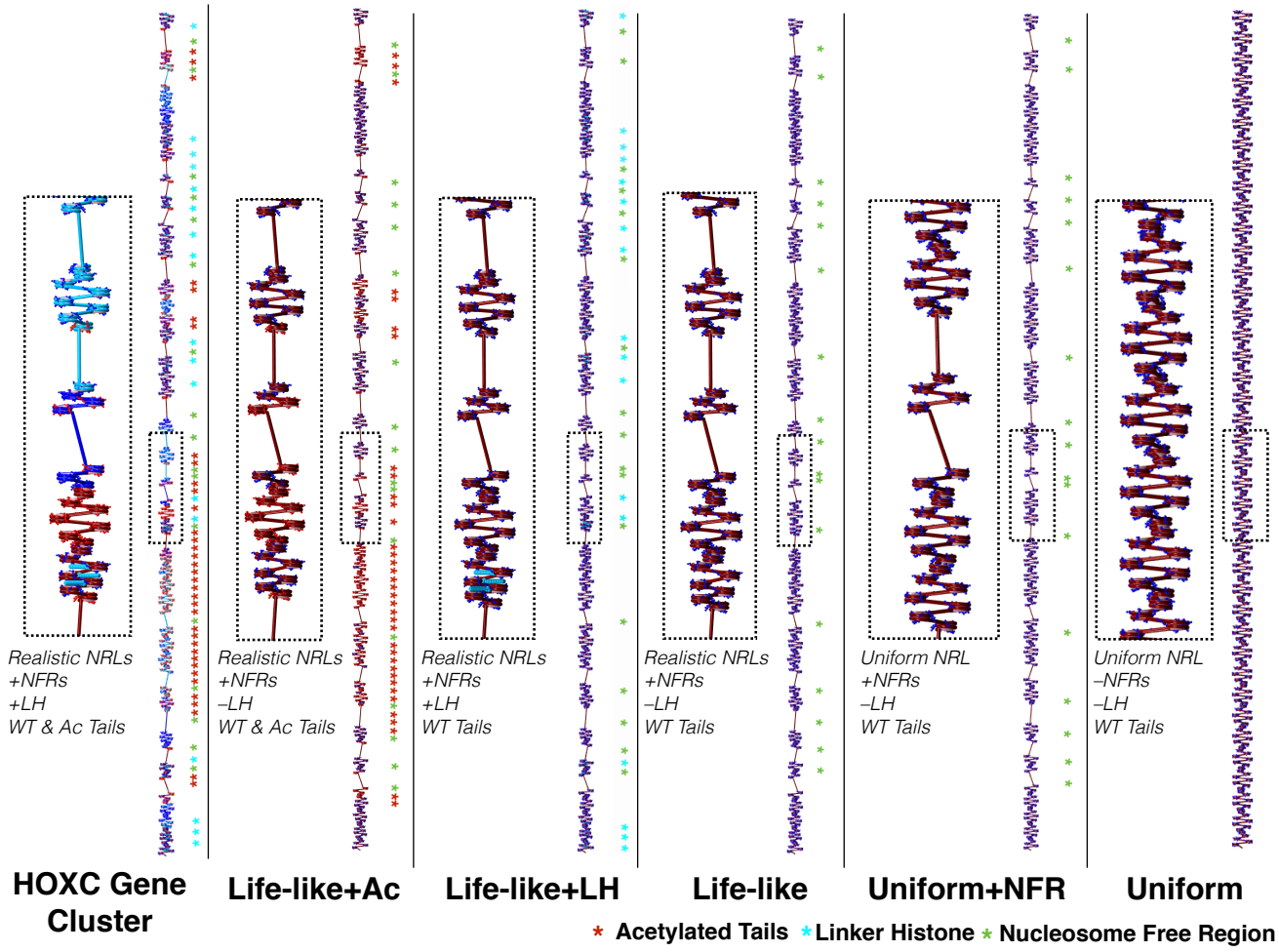

Fig. S1. Starting configurations for all systems studied. From left to right: $\mathbf{HOXC}$, $\mathbf{Life\text{-}like{+}Ac}$, $\mathbf{Life\text{-}Like{+}LH}$, $\mathbf{Life\text{-}Like}$, $\mathbf{Uniform{+}NFR}$ and $\mathbf{Uniform}$ systems respectively.

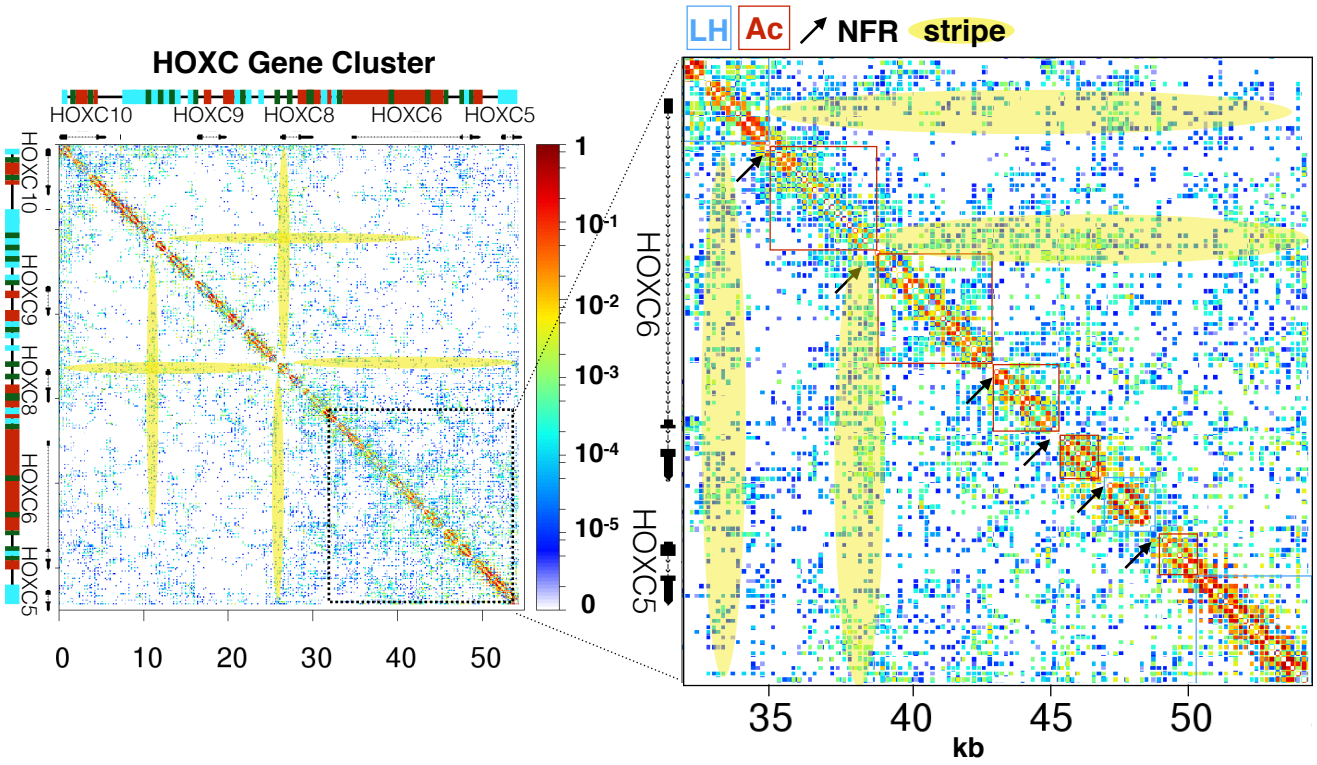

Fig. S2. Close-up view of contact probability maps and resulting fiber structures for $\mathbf{HOXC}$ system. Regions with LH are indicated with a teal square, whereas areas with acetylated histone tails are indicated by a red square. NFR's are indicated with a black arrow, and stripes are highlighted in yellow.

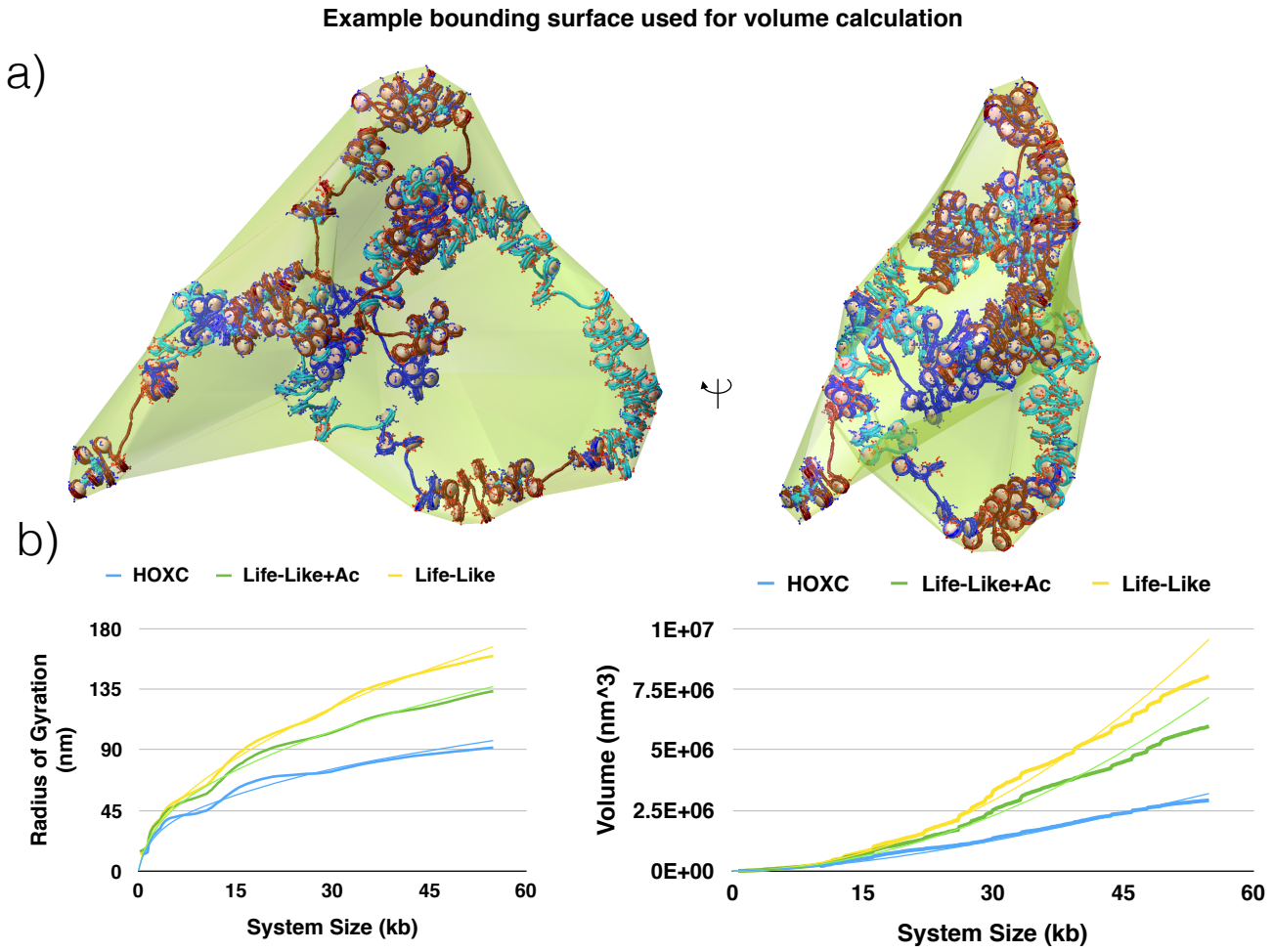

Fig. S3. Rendering of the bounding surface drawn for calculating volume and volume scaling exponent. (a) The bounding surface is solved and plotted using AlphaShape command in matlab, with an alpha parameter of 100, resulting in a convex hull. (b) R_g^2 and Volume as a function of system size, with fitted exponentials used to calculate scaling exponents for $\mathbf{HOXC}$, $\mathbf{Life\text{-}Like{+}Ac}$ and $\mathbf{Life\text{-}like}$ systems.

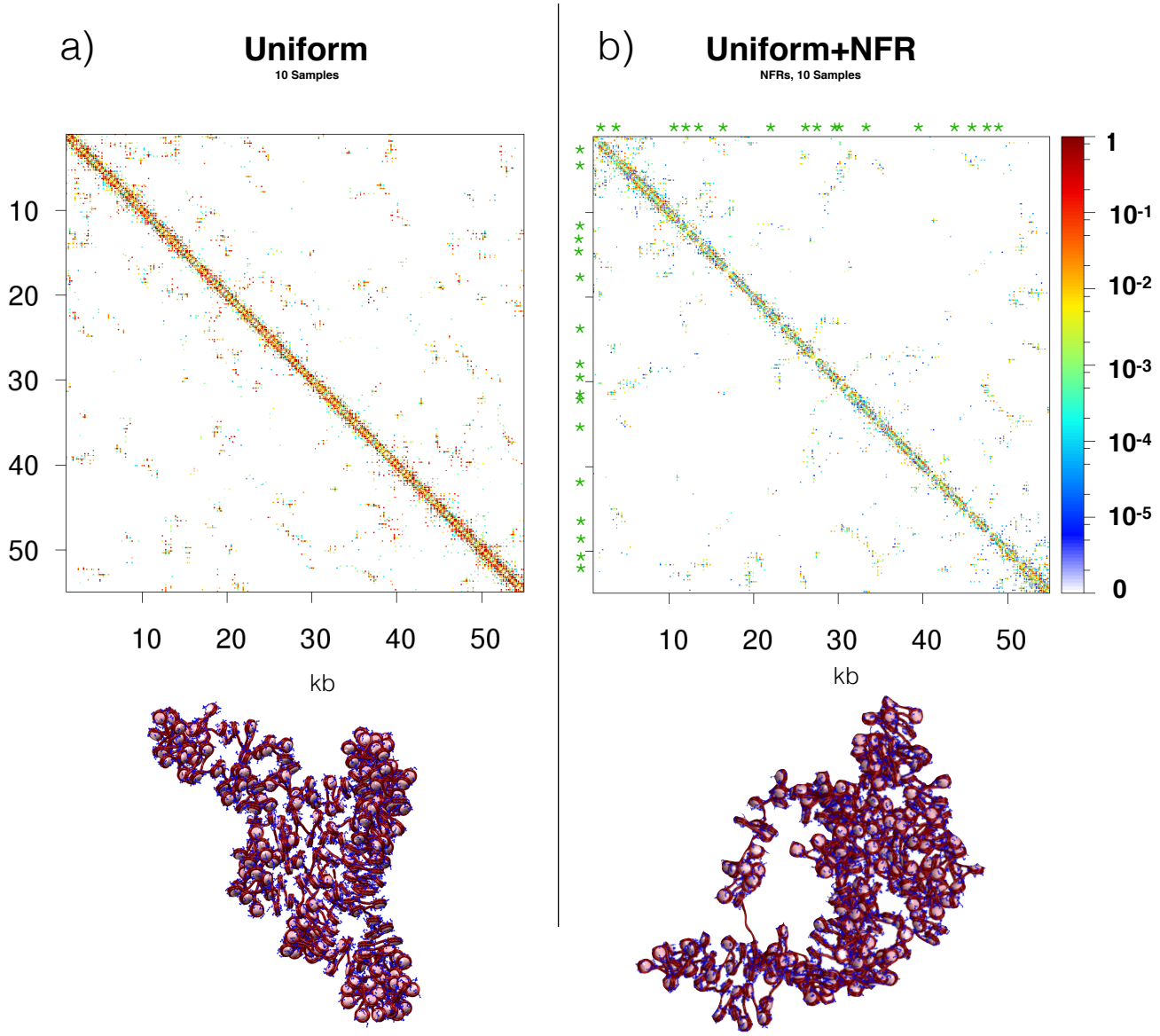

Fig. S4. Fiber renderings and contact probability matrices for (a) $\mathbf{Uniform}$ and (b) $\mathbf{Uniform{+}NFR}$ systems. Green stars indicate the position of NFRs.

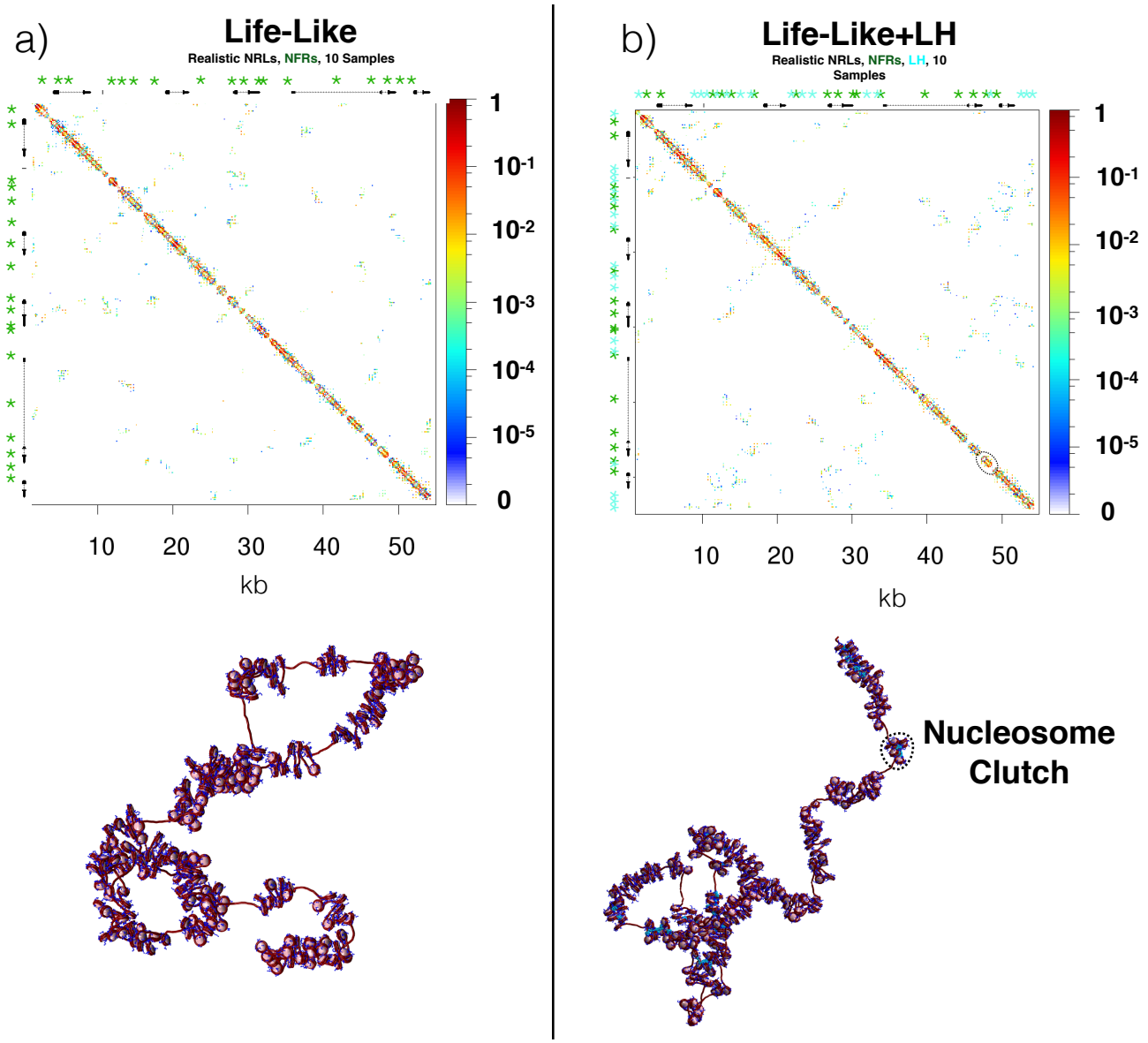

Fig. S5. Fiber renderings and contact probability matrices for (a) $\mathbf{Life\text{-}like}$ and (b) $\mathbf{life\text{-}like{+}LH}$ systems. Black bars along the top of the life-like system show the positions of HOXC genes. Green stars indicate the position of NFRs, and light blue stars indicate LH.

References

1. Luque A, Collepardo-Guevara R, Grigoryev S, Schlick T (2014) Dynamic condensation of linker histone C-terminal domain regulates chromatin structure. *Nucl. Acids Res.* 42:7553–7560.
2. Collepardo-Guevara R, et al. (2015) Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: a multiscale computational study. *J. Amer. Chem. Soc.* 137:10205–10215.
3. Rao SS, et al. (2017) Cohesin loss eliminates all loop domains. *Cell* 171:305–320.
4. Bascom G, Schlick T (2018) Chromatin fiber folding directed by cooperative histone tail acetylation and linker histone binding. *Biophys. J.* 114(10):2376–2385.
5. Sun J, Zhang Q, Schlick T (2005) Electrostatic mechanism of nucleosomal array folding revealed by computer simulation. *Proc. Natl. Acad. Sci. USA* 102:8180—8185.
6. Perišić O, Collepardo-Guevara R, Schlick T (2010) Modeling studies of chromatin fiber structure as a function of DNA linker length. *J. Mol. Biol.* 403:777–802.
7. Boettiger AN, et al. (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529:418–422.